Cuestionario de teoría 2

Antonio Álvarez Caballero
analca3@correo.ugr.es

12 de mayo de 2016

1. Cuestiones

Cuestión 1. Sean x e y dos vectores de observaciones de tamaño N. Sea

$$\mathrm{cov}(x, y) = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})$$

la covarianza de dichos vectores, donde $\bar{z}$ representa el valor medio de los elementos de z. Considere ahora una matriz X cuyas columnas representan vectores de observaciones. La matriz de covarianzas asociada a la matriz X es el conjunto de covarianzas definidas por cada dos de sus vectores columnas. Defina la expresión matricial que expresa la matriz cov(X) en función de la matriz X.

Solución. ss

Cuestión 2. Considerar la matriz hat definida en regresión, $\mathrm{H} = \mathrm{X}(\mathrm{X}^{\mathrm{T}}\mathrm{X})^{-1}\mathrm{X}^{\mathrm{T}}$, donde X es una matriz $N \times (d + 1)$, y $\mathrm{X}^{\mathrm{T}}\mathrm{X}$ es invertible.

1. Mostrar que H es simétrica
2. Mostrar que $\mathrm{H}^{\mathrm{K}} = \mathrm{H}$ para cualquier entero K

Solución. d

Cuestión 3. Resolver el siguiente problema: Encontrar el punto (x_0, y_0) sobre la línea $ax + by + d = 0$ que este más cerca del punto (x_1, y_1).

Consideremos el problema de optimización lineal con restricciones definido por

$$\mathrm{Min}_z \mathrm{c}^{\mathrm{T}}\mathrm{z}$$
$$\text{Sujeto a } \mathrm{Az} \leq \mathrm{b}$$

donde c y b son vectores y A es una matriz.

1. Para un conjunto de datos linealmente separable mostrar que para algún w se debe de verificar la condición $\mathrm{y_n w^T x_n} > 0$ para todo $(\mathrm{x_n}, \mathrm{y_n})$ del conjunto.
2. Formular un problema de programación lineal que resuelva el problema de la búsqueda del hiperplano separador. Es decir, identifique quienes son A, **z**, **b** y **c** para este caso.

Solución. d

Cuestión 4. Probar que en el caso general de funciones con ruido se verifica que $\mathbb{E}_{\mathcal{D}}[E_{out}] = \sigma^2 + \texttt{bias} + \texttt{var}$ (ver transparencias de clase).

Solución. cd

Cuestión 5.

Consideremos las mismas condiciones generales del enunciado del Ejercicio.2 del apartado de Regresión de la relación de ejercicios.2. Considerar ahora $\sigma = 0{,}1$ y $d = 8$, ¿cual es el más pequeño tamaño muestral que resultará en un valor esperado de E_{in} mayor de 0,008?.

Solución. b

Cuestión 6. En regresión logística mostrar que

$$\nabla E_{in}(w) = -\frac{1}{N}\sum_{n=1}^{N}\frac{y_n x_n}{1+e^{y_n w^T x_n}} = \frac{1}{N}\sum_{n=1}^{N} -y_n x_n \sigma(-y_n w^T x_n)$$

Argumentar que un ejemplo mal clasificado contribuye al gradiente más que un ejemplo bien clasificado.

Solución. c

Cuestión 7. Definamos el error en un punto (x_n, y_n) por

$$e_n(w) = \max(0, -y_n w^T x_n)$$

Argumentar que el algoritmo PLA puede interpretarse como SGD sobre e_n con tasa de aprendizaje $\nu = 1$.

Solución. asdffg

Cuestión 8. El ruido determinista depende de $\mathcal{H}$, ya que algunos modelos aproximan mejor f que otros.

1. Suponer que $\mathcal{H}$ es fija y que incrementamos la complejidad de f.
2. Suponer que f es fija y decrementamos la complejidad de $\mathcal{H}$

Contestar para ambos escenarios: ¿En general subirá o bajará el ruido determinista? ¿La tendencia a sobrejaustar será mayor o menor? (Ayuda: analizar los detalles que influencian el sobreajuste)

Solución. dfasdf

Cuestión 9. La técnica de regularización de Tikhonov es bastante general al usar la condición

$$wt\Gamma^{\mathrm{T}}\Gamma w \leq C$$

que define relaciones entre las w_i (La matriz Γ_i se denomina regularizador de Tikhonov)

1. Calcular Γ cuando $\sum_{q=0}^{Q} w_q^2 \leq C$
2. Calcular Γ cuando $(\sum_{q=0}^{Q} w_q)^2 \leq C$

Argumentar si el estudio de los regularizadores de Tikhonov puede hacerse a través de las propiedades algebraicas de las matrices Γ.

Solución. asdf

BONUS. Considerar la matriz $\hat{\mathrm{H}} = \mathrm{X}(\mathrm{X}^{\mathrm{T}}\mathrm{X})^{-1}\mathrm{X}^{\mathrm{T}}$. Sea X una matriz $N \times (d+1)$, y $\mathrm{X}^{\mathrm{T}}\mathrm{X}$ invertible. Mostrar que traza(H) $= d+1$, donde traza significa la suma de los elementos de la diagonal principal. (+1 punto)